Assessing the Generalizability of Integrating Speech Across Space

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in Psychology in the undergraduate colleges of The Ohio State University

by

Zhixin Zhou

The Ohio State University

April 2021

Project Advisor: Professor Mark Pitt, Department of Psychology

Abstract

In a crowded room, listeners easily organize and understand speech arising from various sources. This process involves auditory organization, specifically requiring listeners to organize spoken words coming from different locations. To investigate this, Freggens and Pitt (in preparation) created a cross-ear integration paradigm to measure the organization of speech split in location. Participants hear words either intact (span) or separated (/s/ on the left and “ban” on the right) and are told to only report what they hear on one side. If integrated across ears, participants should report the word with /s/ attached to the voiceless stop /p t/ (“span”) and, if not, report a word with the voiced stop /b d/ “ban” (span – /s/ = ban). Freggens and Pitt found that even when parts of the word were separated by the maximum distance (180 degrees), participants always integrated the word pieces (“span”). Experiments 1 (N=25) and 2 (N=23) were designed to test the generalization of this phenomenon to other word clusters and other scenes, using spelling change as an indication of integration. Experiment 1 used words starting with the clusters /sp st/ to replicate the findings of Freggens and Pitt as well as these clusters /sl sm sn sw shr fl fr/ to generalize results. I found overwhelming integration for all clusters tested, suggesting default integration for speech. Experiment 2 was conducted to investigate whether default integration occurs with more complex speech scenes. Participants performed the same task but heard two simultaneous words (a target word and a competing word) instead of one split word. The first phoneme of the competing word served the same function as the isolated fricative from Experiment 1, and spelling-change was used to indicate integration. Unlike in Experiment 1, I found a very low integration rate for all clusters in the experiment. This means that the competing word on the unattended side interrupted the default integration across locations found in Experiment 1. In conclusion, Experiment 1 supports the default integration mechanism for

speech organization, although Experiment 2 extends the literature by showing that this integration mechanism for speech organization is limited only to situations where the speech sound has nothing else to attach with.

Assessing the Generalizability of Integrating Speech Across Space

Listeners hear sounds from various locations and sources and easily organize them into perceptual objects. For example, in a crowded room full of speech and sounds, listeners can easily distinguish their friends' speech from others. All those sounds reach the ears as one averaged signal, so no organization is inherent in what the listeners hear. However, since listeners easily disentangle the acoustic mixture and extract important messages from it, they must construct the organization. How do listeners do this when objects making sounds are separated in location? This is an important question because listeners often hear complex auditory scenes, which means that people often have to organize sounds coming from various directions and sources. Therefore, this project investigates auditory perceptual organization, specifically how listeners organize spoken words presented to different locations in space.

Similar issues in vision had been discussed by Köhler in Gestalt School of Psychology (Köhler, 1947). He tried to explain why the whole object people visually perceive is sometimes more than the sum of many smaller objects. For example, when a circle has many gaps in it, people perceptually fill in the gaps and still perceive it as a complete circle (principle of closure). Köhler proposed several principles for how individual objects are grouped into a larger picture, such as proximity, similarity, closure, and continuity. All these principles explain the issue from a bottom-up perspective, automatically affecting listeners' perception before higher-level cognitive processing is applied. By incorporating these established visual Gestalt principles, Bregman (1990) proposed auditory scene analysis (auditory perceptual organization), which is the process of organizing mixed auditory information into meaningful groups. In other words, it describes how people separate out sounds from different sources and group acoustics from the same sound source together. Auditory perceptual organization involves both bottom-up (location, frequency, etc.) and top-down processing (schema-based organization). For example, when listening to a spoken sentence, localizing the talker is due to the reliance on spatial location of bottom-up processing and extracting the meaning is due to lexical knowledge of top-down processing.

Bregman (1990) proposed that acoustic similarity and continuity were two important grouping cues. *Continuity* means that uninterrupted auditory objects are more likely to be grouped together, and sudden discontinuation often signals a new auditory object. *Similarity* includes the frequency, timbre, onset time, and location. According to Bregman, auditory objects that perceptibly have the same frequencies, share the same timbre, start at the same time, or emanate from the same place are often grouped together. For example, a harmonic complex tone is made up of several harmonics (acoustic frequencies) that are added at a fixed interval from the original tone. Listeners often perceive a harmonic complex as one sound, but when one of the tones is slightly mistuned, they perceive that mistuned tone as a separate sound from the harmonic (Moore et al., 1986).

Another strong grouping cue that Bregman (1990) proposed was spatial separation, meaning auditory objects that are emanated from different locations tend to be perceived as segregated. With a set of three tone-rating experiments, Barsz (1991) found that spatial separation was a more reliable grouping cue for discriminating tone sequences in comparison to acoustic frequency. When a sequence of sine tones was presented from the same location, participants rated the tones as an integrated complex tone. However, when half of the tones were played from one location and the other half from another location, participants perceived them as two separate tones. In summation, listeners tend to group objects together when they are from the same location and separate them when they are from perceptibly different locations.

Most of the studies on the effect of spatial location as a grouping cue used pure tones as stimuli (Barz, 1991; Hartmann & Johnson, 1991; van Noorden, 1975), but the sounds listeners hear are typically more complex than pure tones. Listeners perceive and produce speech all the time, and they are skilled at grouping mixed spoken phonemes (speech sounds) into meaningful auditory objects. However, many spoken words and even segments of words (phonemes) are acoustically dissimilar from each other and discontinuous (see Jongman et al., 2000 for how each fricative is acoustically different from each other; see Halle et al., 1957 for the silence with producing stop consonants). For example, figure 1 is a spectrogram of the word “span,” with time on the x axis, frequency on the y axis, and the darkness representing intensity. This spectrogram displays the acoustic frequencies of each phoneme /s, p, a, n/ of the word “span.”

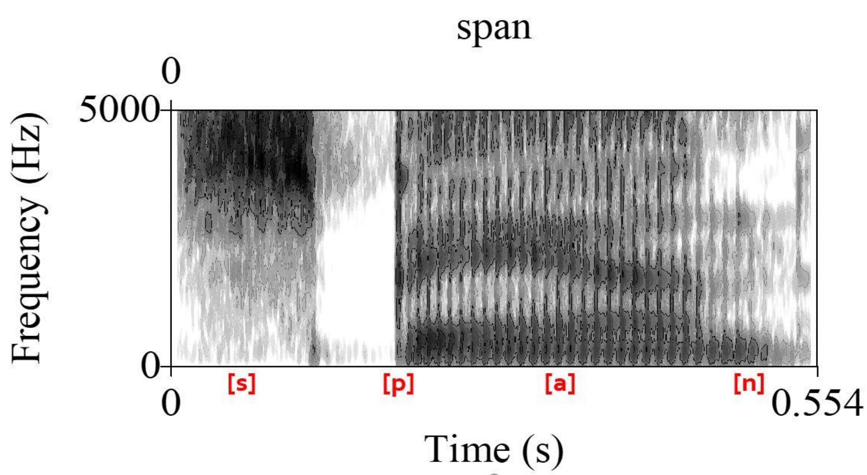

Figure 1 is a spectrogram of the spoken word "span." The x-axis is time (in seconds), and the y-axis is acoustic frequency (in Hertz). Each phoneme making up "span" is indicated in red at the point in time where it begins.

As demonstrated by figure 1, each phoneme in “span” has different active frequencies (dissimilarity), and there is a period of silence before /p/ (discontinuity). This figure shows that the phonemes are dissimilar and discontinuous, and yet listeners easily organize them together into a complete word. In fact, all words with a stop consonant /p b t d k g/ in the middle or the end will display discontinuation within the word. Even though many words are not continuous in the middle, listeners still perceive each word as an integrated whole. For example, when a speaker says “store,” listeners never perceive it as /s/ + “door.” It appears that speech is indeed a little mysterious because it does not strictly follow Bregman’s principles. Since speech is so different from pure tones, it is possible that it does not follow the rule of spatial separation as a grouping cue as well.

Indeed, some scholars found evidence that speech does not adhere to Bregman’s proposal on spatial separation as a grouping cue. Studies on phonological fusion have demonstrated people’s ability to perceptually combine (integrate) words that are separated in space (Cutting, 1975, 1976; Cutting & Day, 1975; Poltrock & Hunt, 1977; Sexton & Geffen, 1981). Fusion occurs when two separated words (“pay” in the left ear and “lay” in the right ear) are perceived as a more complex, blended word (“play”). In Cutting’s experiment, participants were asked to attend to one side while two very similar words (“pay” and “lay”) were presented, one to each side. Cutting found about 50% of participants reported hearing the combined (integrated) word “play.” If people indeed perceive sounds from different locations as distinct auditory objects, participants from Cutting’s study should all report “pay” or “lay” instead of the integrated “play.” Therefore, phonological fusion provides evidence that spatial separation does not necessarily stop listeners from grouping auditory objects, specifically spoken words, together, which challenged Bregman’s (1990) proposal that spatial separation is a strong grouping cue for sounds.

Most phonological fusion studies examined the integration between two separated, complete English words. To investigate whether integration could occur both between and within words, Freggens and Pitt (in preparation) created a cross-ear integration paradigm to measure the organization of speech split in location (“span” into /s/ + “ban”). Similar to Cutting’s (1975) experiment, participants were asked to attend to either their left or right side. They heard words either intact (“span”) or split (/s/ on the left and “ban” on the right) and were told to report only what they heard on their attended side. If integrated across ears, participants should report the word with /s/ attached to the voiceless stop /p t/ (“span”) and, if not, report a word with the voiced stop /b d/ “ban” (“span” – /s/ = “ban”). The sound change (p -> d, t -> d) was an important indicator of integration. In English, when a voiceless stop consonant /p t k/ combines with /s/ to create an initial cluster /sp st sk/, the voiceless stop consonant changes its pronunciation to its voiced counterpart /b d g/. When heard in isolation (without the preceding /s/), English listeners perceive the voiced sound (/b d g/), rather than the voiceless sound (/p d g/). If participants did not perceptually integrate the two parts, they should report hearing “ban” instead of the integrated “span.” Freggens and Pitt found that even when parts of the word were separated by the maximum distance in space (180 degrees), participants always integrated the word pieces (“span”). The finding offered another piece of evidence that people could integrate speech across ears despite spatial separation. As a result of the consistent cross-ear integration with /sp/ and /st/ initial words across all individuals, Freggens and Pitt suggested the possibility of a default integration system for organizing speech. They proposed that during auditory perceptual organization, instead of segregating speech based on where it was emanated, people might be predisposed to automatically integrate speech parts that are separated in space.

Current Study

To be more representative of spoken words in English, this project attempts to generalize results to words with other common initial clusters (Denes, 1963). Participants were asked to attend to one side and type what they heard. Since other phonemes do not resemble the phonotactic property of /p/ and /t/, instead of relying on a change in phoneme (p -> b), this paradigm relied on spelling change (sleep -> /s/ + “leap”). For example, when /s/ was presented to the right and “leap” was presented to the left (attended side), if cross-ear integration occurred, then participants should perceive “sleep”. If the integration did not occur, participants should perceive “leap” instead. The paradigm for Experiment 1 was similar to Freggens and Pitt’s (in preparation) experiment. Critically, the stimuli included /sp st/ initial words to replicate the finding and /fl fr shr sl sm sn sw/ initial words to generalize the finding further. These generalized consonant clusters are acoustically similar to /sp st/ because, like /sp st/, they are all fricative-initial clusters. A fricative is produced when a rapid airflow goes through the constricted oral cavity, and it is characterized by turbulence in the flow (Jongman et al., 2000). Since Freggens and Pitt found overwhelming evidence of integration with /sp st/, I predict that the same high integration rate will also generalize to other acoustically similar word-initial phoneme clusters.

Experiment 2 attempted to replicate Freggens and Pitt’s and Experiment 1’s findings with fricative + stop (“span”) and fricative + nasal/liquid (“sleep”) clusters and further generalize the finding to other word initial clusters common to English to investigate if the integration mechanism (Freggens & Pitt, in preparation) can be applied to all English phoneme clusters, including fricative + vowel (“fake”), nasal/liquid + vowel (“nap”), stop + vowel (“pie”), and stop liquid (“blown”). To accommodate more diverse clusters, the task differed slightly with the presence of a competing word instead of an isolated phoneme (which will be explained in detail later). However, the prediction remains the same that participants should integrate word pieces regardless of the spatial separation based on the assumption that integration should occur for all initial clusters under all circumstances.

If the hypothesis that all phoneme-combination conditions tested should demonstrate a high integration rate is supported, then the data will support the default integration system that fuses speech parts regardless of spatial separation (Freggens & Pitt, in preparation) by expanding the finding to other common word initial clusters of English, which contributes to our understanding of how people organize spoken words in the environment.

Experiment 1

Freggens and Pitt conducted their cross-ear integration experiment with /sp/ and /st/ initial words, and they found participants overwhelmingly integrated the separated speech parts (reporting “spin” instead of “bin”), regardless of the spatial separation. To further generalize the finding, Experiment 1 replicated their /sp/ and /st/ experiment as well as extended the stimuli to other fricative clusters /fl fr shr sl sm sn sw/. These clusters were all combinations of a fricative and a consonant, so they were acoustically similar to /sp/ and /st/, meaning that they should behave similarly. The paradigm remained the same as Freggens and Pitt’s experiment. I predicted that, similar to Freggens and Pitt’s /sp st/ results, participants would integrate the other fricative clusters, even when spatially separated.

Method

Participants

To be eligible for this experiment, participants had to be at least 18 years old. After removing 5 participants’ data for analysis due to being nonnative speakers of English and 6 for failing the fillers check, this experiment included 25 participants (7 attended to the left ear, 18 attended to the right ear). All participants were from an introductory psychology course, and they participated in exchange for course credits at Ohio State University. All participants self-reported normal hearing ability. Demographic data on gender and age were not collected.

Stimuli

A native-English speaking male recorded monosyllabic English words (like “sleep”) to make the stimuli for this experiment. All words were common American English words (See Appendix A). All words of target trials were specifically chosen because they were English words both with or without the first phoneme (“sleep” or “leap”). For example, if the trial was “sleep,” then /s/ was the first phoneme that could be attached to or separated from the base “leap.” In this case, “sleep” and “leap” were both real English words. The average amplitude of all stimuli was normalized to 75dB using a custom Praat script (Boersma & Weenink, 2021). All stimuli were then lateralized using a custom Python script involving Head Related Transfer Functions (Kayser et al., 2009) so that the words were perceived to be presented to the left or the right ear primarily. Lateralization is the process of modifying the intensity of two channels of the sound files to create different intensity levels between the two ears. Due to this lateralization, participants perceived the sound as being presented to the left ear or the right ear, though the sound was present (to different amplitudes) in both ears. Left-lateralized stimuli were perceived as played to the left ear, and right-lateralized stimuli were perceived on the right. When one word (“leap”) was left-lateralized and the first phoneme (/s/) was right-lateralized, participants perceived the two parts as being 180 degree separated in space.

Similar to Freggens and Pitt’s experiment (in preparation), each version of the target stimulus had the first phoneme being either attached to the word on the attended side or presented separately on the unattended side (see figure 2). These were combined such that in every intact trial, one word was played to the attended side, and silence was played to the other side (for example, “sleep” on the left; see figure 2a). In every split trial, one word was perceived to be coming from the attended side (for example, “leap” on the left), and the first phoneme was perceived to be coming from the other side (for example, /s/ on the right; see figure 2b). To examine whether the first phoneme on one side could perceptually integrate with the word on the opposing side, the word on the attended side was started asynchronously after the first phoneme. In other words, even though /s/ was moved to the other side, the temporal alignment between /s/ and the rest of the word was unchanged. For example, if /s/ was played on the right side, then “leap” would be played on the left side 154 milliseconds after the onset of /s/ (the length of /s/). The onset time difference was created to ensure that the first phoneme could perceptually combine with the word. Like Freggens and Pitt’s experiment (in preparation), each trial comprised of a base word (“leap) and a first phoneme (/s/). Base words were always played on the attended side.

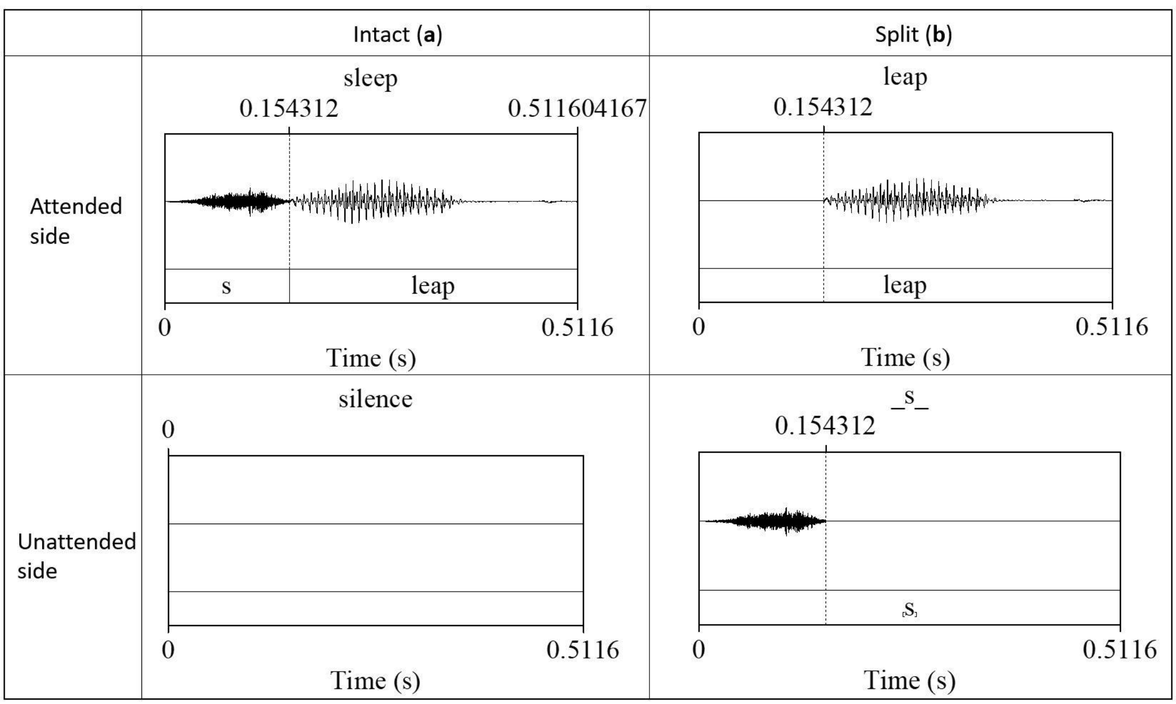

Figure 2. The acoustic waveforms of target trials are shown above, using the word “sleep.” The two columns represent the lateralized condition (intact on the left, split on the right). The rows show the sound presented to each side. 2a demonstrates an intact trial (“sleep” and silence), and 2b is a split trial (“leap” and /s/).

To adequately generalize to other word initial clusters of English, there were nine phoneme-combination conditions: 12% /fl/ (“fleece”), 13% /fr/ (“freight”), 9% /shr/ (“shriek”), 10% /sl/ (“sleep”), 10% /sm/ (“smear”), 10% /sn/ (“snow”), 13% /sp/ (“spark”), 12% /st/ (“store”), and 11% /sw/ (“sweat”). All clusters were fricative-consonant clusters so that they were acoustically similar to /sp/ and /st/ in Freggens and Pitt’s experiment.

Two sets of stimuli were created so that each set consisted of one version of the stimulus pair (intact and split counterbalancing). For example, if “sleep” were the intact trial for one list, then it would be the split trial for the other list. Each set also had two list versions to counterbalance the participants’ attended side (left and right). There were four different stimulus lists, and this experiment had a within-subject repeated design, so each participant was randomly assigned to listen to one of the four lists.

The target trials within each list were interspersed with 40 fillers. The fillers were monosyllabic words that could not be integrated cross ears (“root”) because, without the first phoneme, the fillers are no longer real English words. Including fillers made it less likely for the participants to guess the purpose of the experiment and use strategies, such as simply combining the two parts to complete the experiment, instead of reporting their true perceptions.

Procedure

This experiment was hosted online through a custom-built experimentation framework, and the participants used lab computers and headphones to complete this experiment. The experiment took about 15 minutes to complete on average. Before the start of the experiment, participants went through a short task that examined their ability to distinguish the location of stimuli and to make sure they were wearing headphones. A 500Hz tone was played either on the left or the right five times, and participants were asked to click “Right Ear” if it was played either on the left or the right headphone (via Interaural Timing Differences of 200 microseconds), and participants were asked to click “Right Ear” if it was on the right and “Left Ear” if it was on the left. Participants had to pass this task to enter the experiment (it was five trials long).

Participants typed out what they heard on their attending side. They were given 10 seconds to respond before the program moved onto the next trial. Instructions were presented for participants to inform them of the attended side and their task, which was to listen to the entire word and respond by typing the answer on each trial. Then participants completed 5 practice trials, which were fillers and intact trial stimuli. They were instructed to type the word they heard on their attended side (either right or left, depending on the list). On the screen, they would always see a prompt for which side to attend (“Report the word you hear on the LEFT”) and the box to type their response. Participants were only allowed to listen to the stimuli once, but they could edit their typed response before moving onto the next trial. The participants then started the test session, in which the layout of the experiment and the instructions were exactly the same as the practice session. They went through 2 blocks of test trials. Each block consisted of 60 test trials with 40 target trials (50% intact and 50% split) and 20 fillers. The target trials and fillers were presented randomly for each participant.

After finishing the test trials, participants filled out a questionnaire hosted by Qualtrics (Qualtrics, Provo, UT), which asked for their native language. It also collected information on whether the participants have noticed anything from the experiment. If participants reported using strategies that undermine the goal of the experiment, their data would be removed from the analysis. In this experiment, no participant reported or was found to use strategies.

Results and Discussion

Data cleaning

Data of participants who did not meet the 75% correct rate criterion for filler trials were removed (N = 6). Getting a low correct rate for filler trials meant that participants either failed to follow the instructions or did not pay attention. Therefore, the remaining 25 participants' data were included in the analysis. If the participant's response was the same as the base word ("leap"), then the participant did not integrate the stimuli because the first phoneme and the base were perceived as separate objects (segregated responses). If the response was the base word with the first phoneme intact ("sleep"), then integration occurred because the participant perceptually combined the base and the first phoneme despite spatial separation (integrated responses). Moreover, the spelling must have changed for integrated vs. segregated responses, such that the spelling of /s/ + "leap" would be "sleep" instead of "sleap." Similar to the sound change in Freggens and Pitt's experiment, the spelling change indicates real integration. If participants simply strategically combine the first phoneme and the base, they will report the word with unchanged spelling ("sleap"). There was at least one word with a spelling change under each phoneme-combination condition to check if the participant had used strategies. No participant was removed due to this reason.

Predictions

I predicted that Freggens and Pitt's (in preparation) finding that participants always integrated word pieces that were separated in space could be replicated with /sp/ and /st/ and be generalized to other acoustically similar word initial clusters that were tested in this experiment /fl fr shr sl sm sn sw/. Both the intact trials and the split trials should have a high integration rate, with intact trials being slightly higher.

Hypothesis testing

Figure 3 shows laterality on the x-axis and the proportion of integrated responses on the y-axis, with intact trials in red and split trials in blue. The dots represent each participant's integration rate for each laterality condition. For intact trials, the first phoneme was attached to the base words, and silence was presented to the other side ("sleep" = base). The participants should report hearing the integrated base word because it was the only sound present. For split trials, the first phoneme was separated from the base word, and they were played on opposing sides (/s/ + "leap" = "sleep," "leap" = base, /s/ = first phoneme), so participants were expected to perceive the integrated word ("sleep"). As predicted, the average proportion of integrated responses for intact trials was very close to 100% (*M* = 0.98, *SEM* = 0.004). This makes sense because, for intact trials (red box/dots), participants heard the base with the first phoneme attached to the same side ("sleep"). The average of integrated responses for split trials (blue box) was slightly lower than the intact trials but was still very high overall (*M* = 0.93, *SEM* = 0.027). This supports the hypothesis that participants perceptually integrated the speech parts even when they were separated in space. For split trials, high integrated responses meant that contrary to how the stimuli were presented (first phoneme to the opposing ear), participants perceived the isolated first phoneme with the base on their attended side. The intact trial was the control group, and its integration rate provided a baseline to compare the split trial to. As shown by the error bars, there were few deviations around the mean (SEM) for both conditions. This reflected minimal individual variations on the ability to integrate sounds across locations (this will be discussed later).

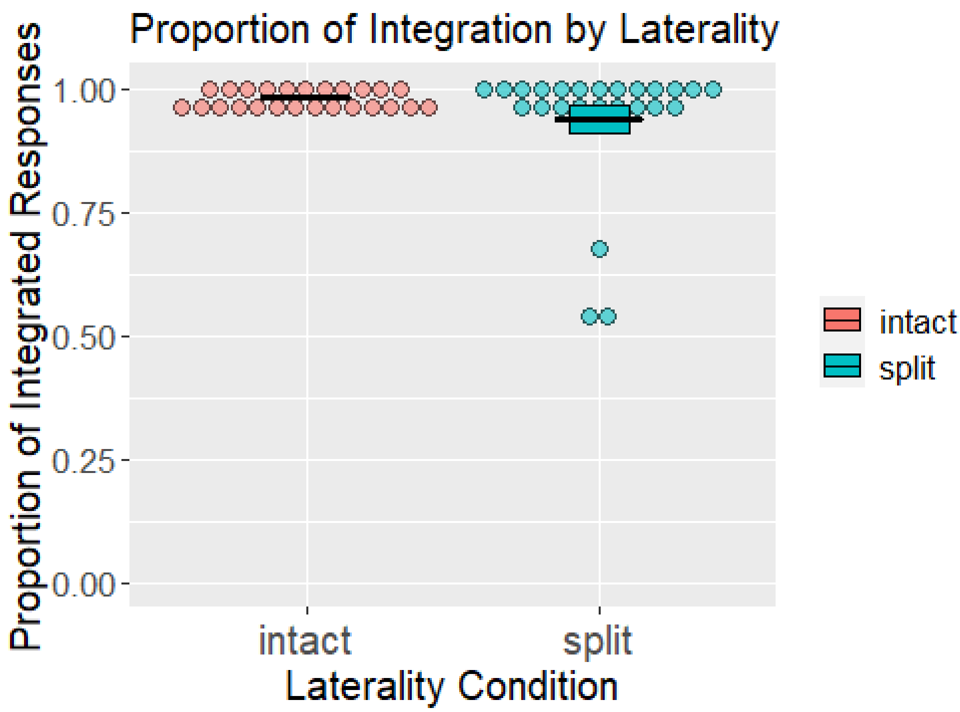

Figure 3. The dot-plot shows laterality conditions (intact and split) on the x-axis and the proportion of integrated responses on the y-axis. The black bars indicate the mean of each laterality, and the upper and lower end of the boxes were the standard error of the mean.

A two-way within-subject repeated ANOVA test was performed to compare the integration rates of intact trials and split trials. The test returned no main effect for laterality [$F(1, 24) = 2.67$, $p = 0.12$]. In other words, the experimental split trial (93%) did not differ statistically from the control intact trial (98%). Therefore, this result supports the hypothesis that participants integrate sounds even though the sounds were completely separated in space.

The purpose of this experiment was to assess the generalizability of integrating speech to acoustically similar phoneme combinations. Figure 4A showed the proportion of integrated responses for phoneme combinations and laterality (split/intact). As shown in figure 4A, not only was there not much difference between intact trials and split trials in any phoneme condition, but it also showed little difference between phoneme conditions for each laterality. This supported the hypothesis that Freggens and Pitt's finding could be generalizable to other word initial clusters that were acoustically similar to /sp/ and /st/.

The difference between conditions did not appear to be large for each laterality. However, the same two-way within-subject repeated ANOVA test found a significant main effect for phoneme-combination conditions [$F(3, 65) = 3.12$, $p = 0.04$]. Post hoc comparisons using t-test with Bonferroni correction indicated that the mean score of /sw/ ($M = 0.91$) was significantly different from that of /sn/ ($M = 0.98$) [$t(49) = 4.07$, $p = 0.01$]. The difference between /sw/ ($M = 0.91$) and /st/ ($M = 0.98$) was also significant [$t(49) = 3.54$, $p = 0.03$]. The /sw/ condition had relatively lower integration rates for both intact and split trials. As a result, their averaged integration rate was statistically significantly lower than the average of /sn/ and /st/ conditions, which were the highest among phoneme-combination conditions. Though this difference between conditions is statistically significant, it is practically small, indicating that participants are 7% less likely to integrate /sw/ across ears than other phonemes. /w/ is a glide consonant, which is also called a semi-vowel, so it is phonetically similar to a vowel. As a result, it is acoustically more different from the other consonants included in this experiment. Participants might be slightly reluctant to attach the first phoneme to a base starting with a more vowel-like consonant. Since /sw/ still had high integration ($M = 0.91$), the difference between intact and split trials was small regardless of the phoneme-combination conditions.

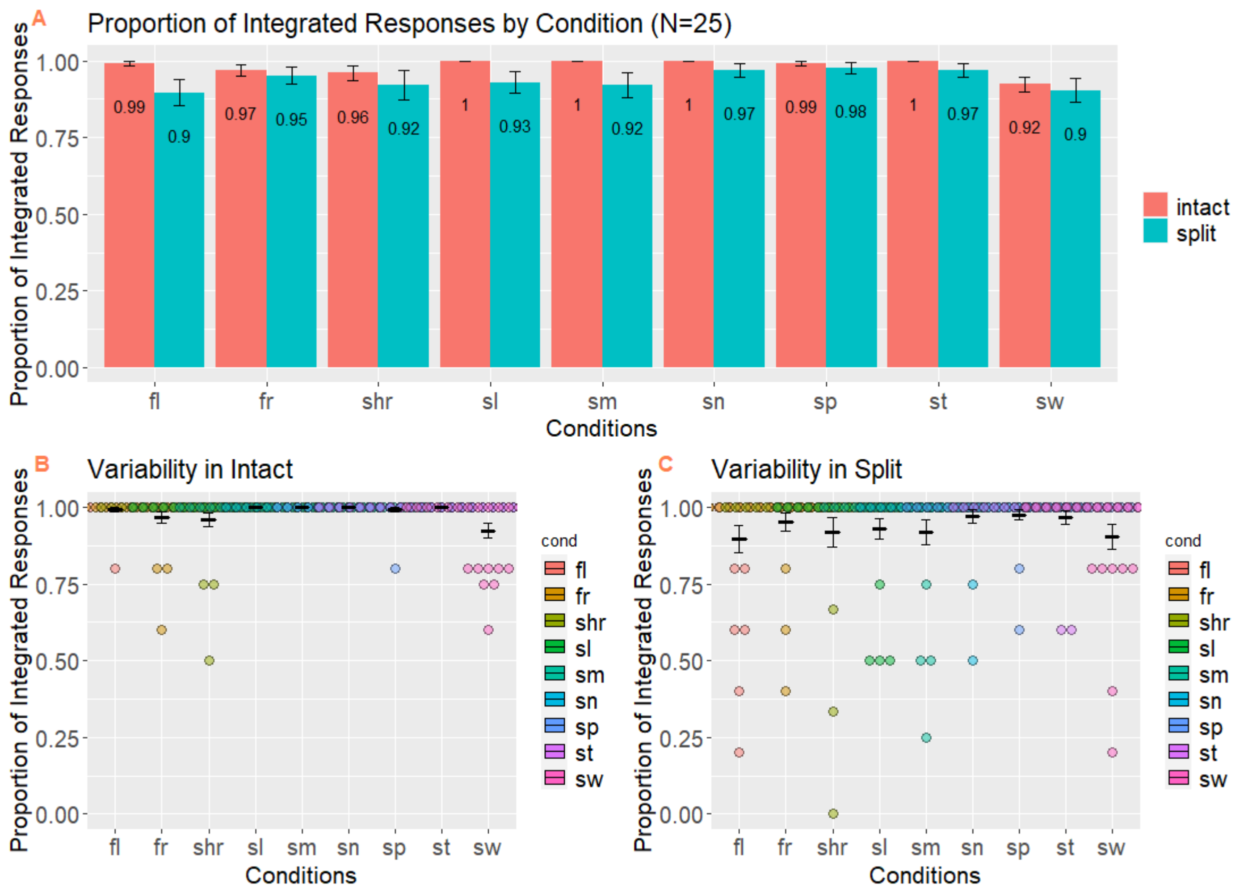

Figure 4 shows the variability of results from Figure 3. Nine phoneme-combination conditions were on the x-axis, and the proportion of integrated responses was on the y-axis. Figure 4a shows the mean performance for each condition, grouped by laterality with intact trials in red and split trials in blue. The number on each bar shows the condition mean, and the error bars showed standard error around the means. Figures 4b and 4c show the variability for each phoneme condition, separated by laterality. Each dot represents a participant within the phoneme condition (each participant has 9 dots), and the error bars showed the means and the standard error around the means.

The ANOVA did not find a significant interaction between phoneme-combination conditions and laterality [$F(4, 99) = 1.62$, $p = 0.17$]. Therefore, the effects of laterality and phoneme-combination conditions were independent. This result was consistent with figure 4A, which displayed an overall high integration rate across phoneme-combination conditions.

Investigation of individual variations

Figures 4B and 4C displayed the variabilities for each condition and laterality, with each dot representing one participant's averaged integration rate for the given phoneme-combination condition. Split trials had larger individual differences than intact trials, which were demonstrated by the distribution of the dots. For intact trials, participants were not required to perceptually integrate stimuli across ears, so the majority of participants typed the correct responses. However, split trials required participants to integrate the stimuli presented on their left and right sides, so I see more individual differences in figure 4C. One possible explanation for this individual variation is that some participants might have better attentional abilities, so they were better able to ignore the stimulus presented on their unattended side. Since all participants included passed the fillers check, they should all be following the instructions. However, the cut-off criterion was 75% of correct rate for fillers. Data of participants who did not pay attention to the whole experiment could have contributed to the variability in the responses.

Summary

The results of Experiment 1 demonstrate a high integration rate for both intact trial and split trials across all phoneme-combination conditions, which support the hypothesis that Freggens and Pitt's finding can generalize to other acoustically similar word initial clusters. This experiment provided more evidence for the potential default system that perceptually integrates spatially separated speech parts.

Experiment 2

Experiment 1 replicated and generalized Freggens and Pitt's (in preparation) findings to other acoustically similar fricative clusters. In this second experiment, I would like to further generalize the finding to other word initial clusters that are acoustically different from the previous ones so that the default integration system can be applied to all English word initial clusters. However, unlike fricatives, other consonants like nasals, liquids, and stops often coarticulate with the vowels following them. So, these consonants were less distinguishable from the rest of the word when isolated. It was difficult to cut the first phoneme off from a word during stimulus creation, and it was also challenging for participants to recognize the first phoneme when presented to one ear alone. To accommodate this issue, in Experiment 2, I used a competing word for the first phoneme to attach to. While intact trials stayed unchanged ("fake" = base), for split trials, the first phoneme was attached to the competing word ("ache" = base, "fact" = competing word, first phoneme = /f/). Similar to Experiment 1, spelling-change was implemented to indicate real integration, and all stimuli had different spellings for segregated ("ache") and integrated responses ("fake"). I still predict that participants would integrate across ears for all the clusters being tested because I assume this integration mechanism is inclusive to all English word initial clusters under all circumstances.

Method

Participants

To be eligible for this experiment, participants had to be at least 18 years old, have access to a computer or laptop, and use either a Firefox or Google Chrome browser. After removing 8 participants' data from analysis due to being nonnative speakers of English and 1 participant for using response strategies, this experiment included 23 participants (15 attended to the left ear, 8 attended to the right ear). All participants were from an introductory psychology course, and they participated in exchange for course credit at Ohio State University. All participants self-reported normal hearing ability. Participants consisted of 12 females and 11 males, aged 18-25 (M = 19.3, SD = 1.6).

Stimuli

The same native-English speaking male recorded pairs of monosyllabic English words (like "ache," "fact") to make the stimuli for this experiment. All words were common American English words (See Appendix B). All words for target trials were specifically chosen because they were English words both with or without the first phoneme ("fact" and "act"). For instance, if the pair consisted of "fact" and "ache," then /f/ was the first phoneme that was combined with either word ("act/fact," "ache/fake"). The lateralization and normalization were the same as Experiment 1, except for half of the trials, one sound was lateralized to the left ("fact") and another to the right ("ache"), so participants perceived two words as separated in space.

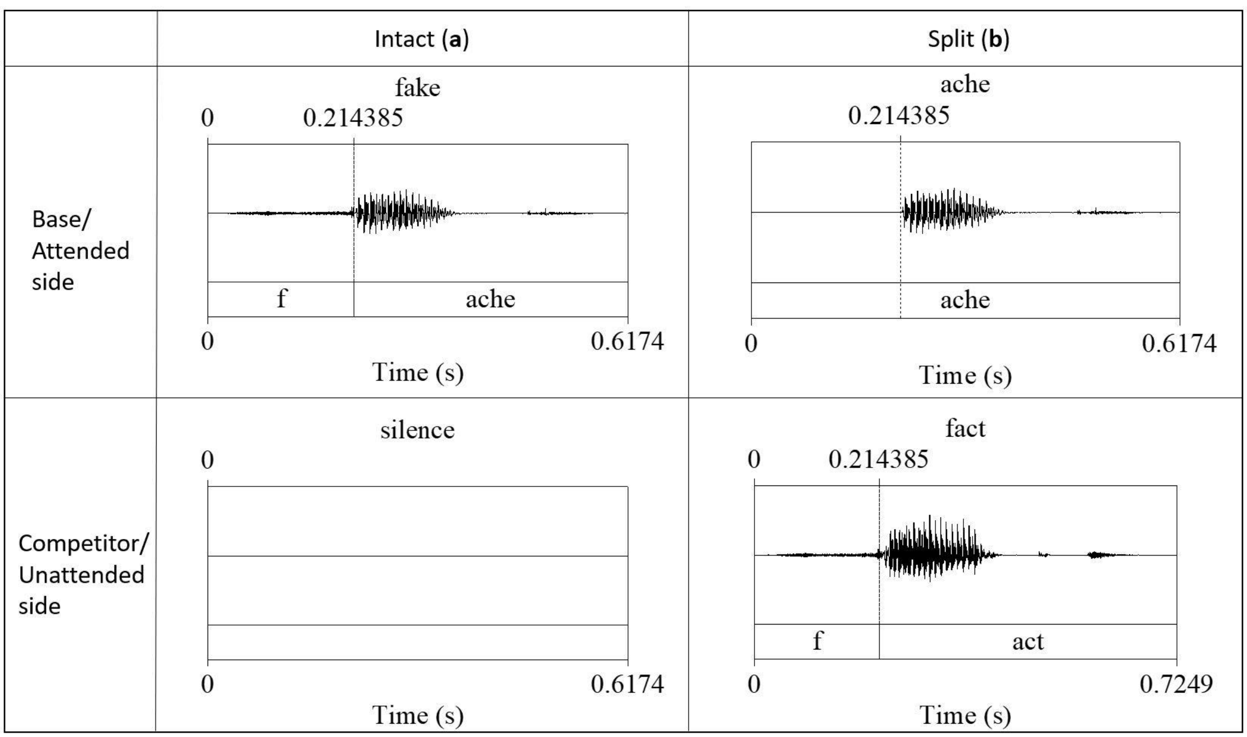

Figure 5. The acoustic waveforms of target trials are shown above, using the word pair "ache" and "fact." The two columns represent the lateralized condition (intact on the left, split on the right). The rows show the sound presented to each side. 5a demonstrates an intact trial ("fake and silence), and 5b is a split trial ("ache" and "fact").

Each version of the target stimuli had the first phoneme (/f/ in "fake") attached to either the attended side or the unattended side (see figure 5). Intact trials remained the same as Experiment 1 (see figure 5a). Unlike Experiment 1, a full competing word was played on the unattended side (for example, "fact" on the right; see figure 5b). The base word was still presented asynchronously after the first phoneme of the competing word to ensure that the first phoneme and the base word could perceptually combine. For example, if "fact" was played to the right, then "ache" would be played to the left 214 milliseconds after the onset of "fact" (the length of /f/ in "fact").

In this experiment, each trial comprised of either a pair of words – one base word ("ache") and one competing word ("fact"), or a single base word ("fake"). Base words still were always played on the attended side. Therefore, intact trials remained the same as Experiment 1 (the first phoneme lateralized with the base), but "split" trials added a competing word to the opposing side for the first phoneme to attach to.

	fricative vowel	**nasal/liquid vowel**	**stop vowel**	**stop liquid**	**fricative stop**	**fricative nasal/liquid**
Consonant Clusters	/f, th, s, sh/	/n, m, l, r/	/p, t , b , d/	/pl, pr, bl, br, gr, kl/	/sp, st/	/fr, fl, sn, shr/
Number of trials per cluster	1, 3, 3, 1	2, 2, 2, 2	1, 3, 2, 2	1, 2, 2, 1, 1, 1	4, 4	2, 3, 2, 1
Types of vowels	/eɪ, i, aɪ, oʊ, ɪ, ɛ, æ, ɔ/	/eɪ, i, aɪ, oʊ, ɛ, æ, ɝ, a/	/eɪ, i, aɪ, oʊ, ɛ, æ, ɔ, ɝ, a, aʊ/	/eɪ, i, aɪ, oʊ, ɪ, æ/	/eɪ, i, aɪ, ɪ, ɛ, æ, ɔ, a/	/eɪ, i, aɪ, oʊ, ɪ, ɛ, æ, a, u, ʌ/

Table 1. The details of each phoneme-combination condition are listed. Each column is a phoneme-combination condition. The first row lists all consonants for each condition. The second row shows the number of trials for each consonant cluster proportional to their frequency in English, totaling eight trials under each condition. The third column lists all the vowels used with each condition.

To adequately generalize to linguistic properties of English (the number of each initial cluster under each phoneme-combination condition is proportionate to its actual frequency of occurrence in speech), trials were categorized into six phoneme-combination conditions (see Table 1): 16.7% fricative vowel (“fake”), 16.7% nasal/liquid vowel (“name”), 16.7% stop vowel (“tone”), 16.7% stop liquid (“blown”), 16.7% fricative stop (“space”), and 16.7% fricative nasal/liquid (“slim”). Within each phoneme-combination condition, the number of each consonant cluster was chosen to match their frequency of occurring in the English language (Denes, 1963). To correct the overabundance of fricative combinations, fillers and practice trials consisted of all stops and nasals/liquid phoneme combinations.

There were still four lists of stimuli created to counterbalance laterality and attending side, and each participant was randomly assigned to listen to one of the four lists. The target trials within each list were interspersed with 40 filler trials. The fillers were pairs of monosyllabic words that could not be integrated across ears (“class” and “moon”) because their combinations are not real English words.

Procedure

The overall procedure and instructions of Experiment 2 were exactly the same as Experiment 1, besides that participants completed the experiment on their own devices (due to the pandemic). After completing 8 practice trials, the participants then started the main experiment. They went through 136 total trials (50% intact and 50% split) with 96 target trials and 40 filler trials. The target trials and fillers were presented completely randomly for each participant.

After finishing the experiment, participants filled out the same questionnaire hosted by Qualtrics (Qualtrics, Provo, UT). One participant reported knowing the purpose of the experiment, and the data were excluded.

Results and Discussion

Data cleaning

Data of one participant who reported using strategies were removed. Everyone else had met the 75% correct rate criterion for fillers. Therefore, the remaining 23 participants' data were included for analysis. The scoring was the same as Experiment 1, and no participant was removed due to falsely reporting words on the unattended side.

Predictions

According to Experiment 1, fricative+stop and fricative+nasal/liquid conditions should have a high integration rate, and the other phoneme-combinations should behave like these two conditions. Although a competing word was present for this experiment, participants should still integrate the first phoneme of the competing word with the base under the assumption that the default integration system (Freggens & Pitt, in preparation) can apply to all initial clusters under all situations. If the integration mechanism is indeed applicable to all word initial clusters under all circumstances, the presence of the competing word should not interrupt integration across ears.

Hypothesis testing

Figure 6 shows the proportion of integrated responses by laterality. Contradicting my prediction, the integration rate for split trials was very low (*M* = 0.04, *SEM* = 0.008; see figure 6). The integration rate for intact trials was fair (*M* = 0.73, *SEM* = 0.02), but it was lower than for Experiment 1 (*M* = 0.98, *SEM* = 0.004). The reason for the lower integration rate for intact trials will be discussed later. Intact trials should ideally have 100% of integrated responses because no competing word was present. The drop in integration rate for intact trials could be due to the stimuli changes, which will also be discussed later. In contrast to intact trials, split trials required participants to perceptually integrate what have been played on their two sides ("fact" + "ache" = "fake"). The extremely low integration rate for split trials indicates the finding of Experiment 1 did not generalize to other clusters. Similar to Experiment 1, if the integration rate of split trials was significantly lower than baseline (control trials), then it confirmed that the hypothesis was not supported.

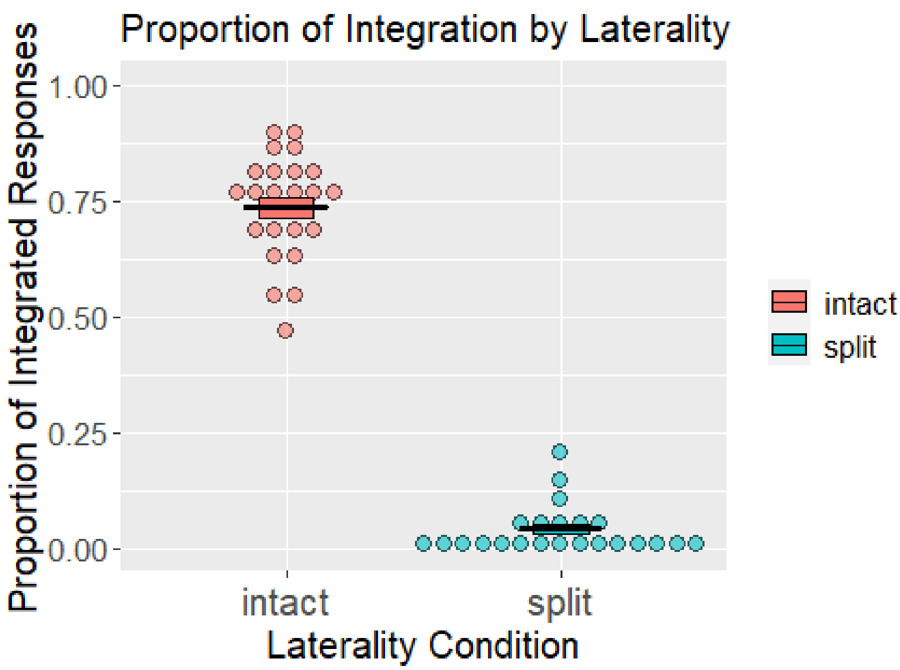

Figure 6. The dot-plot shows laterality (intact and split) on the x-axis and the proportion of integrated responses on the y-axis. The black bars indicate the mean of each laterality, and the upper and lower end of the boxes were the standard error of the mean.

To test if the proportion of integrated response for split trials is significantly lower than the baseline, the same two-way within-subject repeated ANOVA was performed, which found that there is a significant difference between the integration rate of intact and split trials [$F(1, 22) = 909.3$, $p < .001$]. Therefore, the statistical evidence supports that the consistent integration across ears found in Experiment 1 did not occur for the clusters tested in this experiment (fricative+nasal/liquid, fricative+stop, fricative+vowel, nasal/liquid+vowel, stop+liquid, stop+vowel), which contradicts my hypothesis that the integration rate for all the clusters should be high.

The purpose of this experiment was to assess the generalizability of integrating speech to acoustically different phoneme combinations. As shown in figure 7A, not only was there a considerable difference between intact trials and split trials in any phoneme condition, but it also showed a significant difference between phoneme-combination conditions. For experimental split trials, frictive+nasal/liquid and nasal/liquid+vowel had the lowest integration rate (1%) and stop+liquid had the highest integration rate (8%). In Experiment 1, participants perceptually integrated the first phoneme on the other side with the base on the attended side (/s/ + "leap" = "sleep") for all phoneme-combination conditions. These phoneme-combination conditions were included in this experiment under fricative+stop (/sp st/, 3%) and fricative+nasal/liquid (/fl fr sl shr sm sn sw/, 1%) conditions. Their low integration rates failed to replicate Experiment 1. Since Experiment 1 had found that participants were able to integrate speech parts under those phoneme-combination conditions, fricative+stop and fricative+nasal/liquid should have had a high integration rate. Therefore, I assume that the paradigm change (the presence of a competing word) had caused the overall low integration rate. However, no statistical evidence is present to support this hypothesis, so future research is needed.

The same two-way within-subject repeated ANOVA returned a significant main effect for phoneme-combination conditions [$F(3, 60) = 8.3$, $p < .001$] as well as a significant interaction between laterality and phoneme-combination conditions [$F(3, 63) = 17.23$, $p < .001$]. The effect of laterality was analyzed at each phoneme-combination condition with Post hoc comparisons using t-test with Bonferroni correction. These pairwise comparisons indicated significant differences between intact trials and split trials for all phoneme-combination conditions (*p*

< .001), which indicates that all clusters failed to support my hypothesis that participants should integrate the speech parts regardless of the presence of the competing word. The effect of phoneme-combination conditions at each laterality was also analyzed with the same Post hoc test. The test found little difference across phoneme-combination conditions for split trials and many for intact trials (see table 2 for the summarized results). Consistent with figures 7B and 7C, intact trials display larger variability overall than split trials, but unlike split trials which have low variability across all conditions, for intact trials, fricative+nasal/liquid and fricative+stop display significantly less variability. Therefore, the interaction indicates that individual’s performance on integration was not only influenced by the laterality but also the different phoneme-combination conditions, so there was a mixed effect of both factors.

Group 1	Mean	Group 2	Mean	t-Statistics	P-value
fricative+stop	0.94	fricative+vowel	0.75	4.62	0.002
fricative+stop	0.94	stop+vowel	0.60	6.2	< .001
fricative+stop	0.94	stop+liquid	0.62	4.83	0.003
fricative+stop	0.94	nasal/liquid+vowel	0.63	8.24	< .001
fricative+nasal/liquid	0.87	stop+vowel	0.60	4.86	0.001
fricative+nasal/liquid	0.87	stop+liquid	0.62	3.44	0.035
fricative+nasal/liquid	0.87	nasal/liquid+vowel	0.63	5.74	< .001
fricative+vowel	0.75	nasal/liquid+vowel	0.63	3.3	0.049

Table 2 summarizes the significant comparisons found by the post hoc test. They are all under intact trials. Each row shows one comparison. There are eight significant comparisons.

These results were also consistent with figure 7A since fricative+stop and fricative+nasal/liquid indeed had the highest integration rate, and fricative+vowel was also relatively high. Fricative-initial consonant clusters had the highest overall integration rate, followed by the fricative-initial cluster and then the other conditions. Among the other conditions, non-fricative consonant-vowel combinations had the lowest integration rate overall. Though this difference between conditions is statistically significant, it is practically small, indicating that participants are 17% more likely to integrate fricative+stop across ears than other phonemes. The difference between consonant-vowel clusters (“fake”) and consonant-consonant clusters (“freight”) could be ascribed to their difference in articulation. Participants appeared to be more prone to attach an isolated phoneme (consonant) to another consonant instead of a vowel (see table 2). In other word, people might be sightly reluctant to attach the isolated phoneme to the base with an acoustically different first phoneme.

As shown in table 2 and figure 7A, the difference between fricative-initial clusters and other clusters was mostly due to the difference in intact trials (range from 0.60 to 0.94). So, it is possible that this difference could be attributed to the stimuli themselves. During stimulus creation, the male speaker only recorded stimuli for split trials. The stimuli for intact trials were made by lateralizing the first phoneme of the competing word to the base. For example, the pair “ache” (base) and “fact” (competing word) were recorded, and they were the stimuli for the split trial. In order to create “fake” (/f/ + “ache”) for the intact trial, I manually determined the cut-off timing for /f/. It was a lot harder to determine the timing for stops (“take”), liquids (“loan”), and nasals (“nap”) because they coarticulate with succeeding vowels. Many participants reported that the first phonemes were difficult to identify, which was consistent with the overall lower integration rate for intact trials in comparison to Experiment 1.

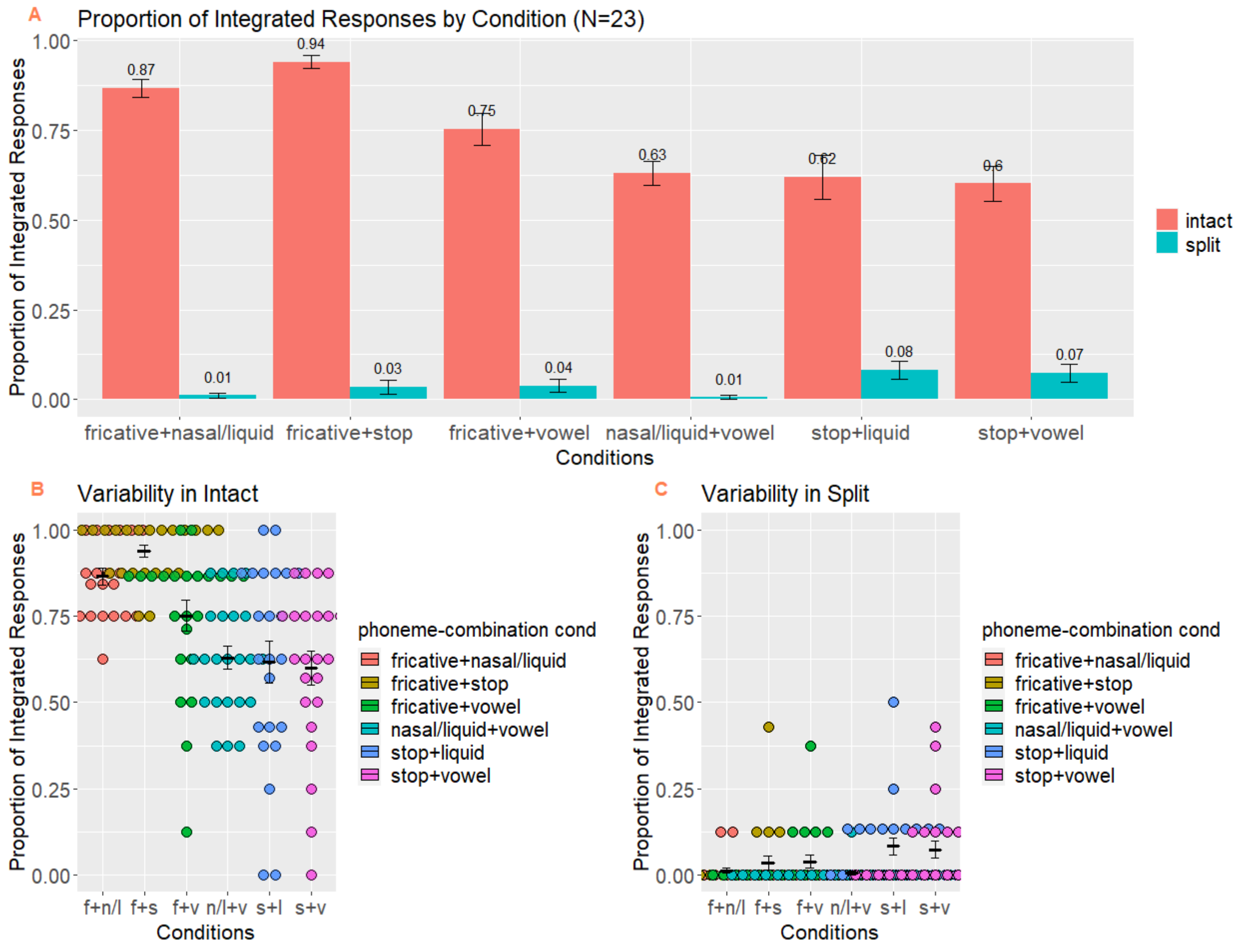

Figure 7 shows the variability of results from figure 6. Six phoneme-combination conditions were on the x-axis, and the proportion of integrated responses was on the y-axis. Figure 7a shows the performance for each condition, grouped by laterality. The number on each bar shows the condition mean, and the error bars showed standard error around the means. Figures 7b and 7c show the variability for each phoneme condition, separated by laterality. Each dot represents a participant within the phoneme condition (each participant has 6 dots), and the error bars showed the means and the standard error around the means.

Investigation of individual variations

As demonstrated by the distribution of dots in figures 7B and 7C, intact trials had larger individual differences than split trials. Split trials required participants to integrate the stimuli presented on their left and right sides, and most participants failed to integrate the two parts, so I see minimal individual differences in figure 7C. Such an overall low integration rate for split trials provides another piece of evidence that when a competing word was present, integration was disrupted. For intact trials, participants were not required to perceptually integrate stimuli across ears, so their responses should have had little variation. As explained above, the performance of intact trials might have been affected by the stimuli. So, some participants might be better at detecting those isolated phonemes, which could be influenced by their own devices, volume levels, and hearing abilities. Also, similar to Experiment 1, data of participants who did not pay attention to the whole experiment could have contributed to the variability in the responses.

Summary

The data demonstrated the low integration rate for split trials across all phoneme-combination conditions. The results did not support my hypothesis, so I could not generalize the finding of Experiment 1 and Freggens and Pitt's experiment to acoustically different word initial clusters. However, I surprisingly found that the presence of a competing word probably caused the drastic drop in integration rate, so I could not conclude that these phoneme-combination conditions used in this experiment will not integrate across space. The data also showed some differences across phoneme-combination conditions for intact trials, which should be fixed in future studies.

General Discussion

Freggens and Pitt's (in preparation) cross-ear experiment found overwhelming integration for /sp/ and /st/ initial words, so they suggested a potential default integration system that automatically integrates word pieces that are separated in space. This project aimed to generalize the finding to other word initial clusters in American English. Experiment 1 provided preliminary support (*M* = 0.93) for the default integration system by replicating Freggens and Pitt's /sp st/ findings, as well as expanding the results to other acoustically similar word-initial clusters /fl fr shr sl sm sn sw/. To further generalize the finding, Experiment 2 implemented a slight paradigm change so that a competing word was played on the unattended side instead of an isolated phoneme. In contrast to my hypothesis, there was a surprisingly low integration rate (M = 0.04) for all phoneme-combination conditions tested (fricative+vowel, fricative+nasal/liquid, fricative+stop, stop+liquid, stop+vowel, nasal/liquid+vowel). In comparison to Experiment 1, Experiment 2 included more acoustically different word initial clusters and implemented a slightly different paradigm. Given that fricative+stop and fricative+nasal/liquid failed to replicate the results found in Experiment 1, the paradigm change (presence of a competing word) may have caused the decrease in integration. Therefore, for now, I cannot conclude that the finding in Freggens and Pitt's experiment and Experiment 1 is not generalizable to other acoustically different word initial clusters. However, the unexpected result discovers a limitation of the default integration system because integration only occurs when a competing word is absent. Experiment 1's results support the proposed default integration mechanism for speech organization. Experiment 2 extends the literature by showing that this integration mechanism for speech organization is limited only to situations where the speech sound has nothing else to attach to.

Although Bregman (1990) laid the theoretical foundation for understanding the auditory perceptual organization, evidence was found that speech violates some of the principles. For example, a word can be perceived as an integrated whole even though there are discontinuation and dissimilarity between phonemes. Bregman proposed that spatial separation is one of the powerful cues for grouping, so when two sounds are presented separately in space, people have the tendency to segregate the sounds. Although Bregman and other scholars found evidence supporting this principle with pure tones (Bregman, 1990; Hartmann & Johnson, 1991; van Noorden, 1975), it is questionable if speech follows this rule. Cutting (1975) conducted an experiment on phonological fusion and found two very similar words can be integrated perceptually despite their spatial separation, which provided a counterexample for Bregman's proposal.

While previous research in phonological fusion (Cutting, 1975, 1976; Cutting & Day, 1975; Poltrock & Hunt, 1977; Sexton & Geffen, 1981) focused on the integration of full words, results of Experiment 1 and Experiment 2, along with Freggens and Pitt's finding (in preparation), expanded such integration to word pieces (/s/ + "leap"). Prior to Freggens and Pitt and this project, no scholars have studied the integration of word pieces across space to my knowledge. Freggens and Pitt's finding suggested that speech violates Bregman's (1990) principle of spatial separation as a strong grouping cue not only between words but within words, which necessitates a different mechanism of perceptually organizing speech in space. Therefore, they proposed a default integration system to explain how listeners perceptually organize speech. The results of this project provide more evidence contradicting Bregman's proposal. More importantly, this project generalizes Freggens and Pitt's finding to other common word initial clusters of English (Dene, 1963), so the results further support Freggens and Pitt's hypothesis of a default integration system that perceptually integrates speech parts that are separated in space automatically. The unexpected result in Experiment 2 provides a valuable insight into a limitation of the proposed default integration system, so it only integrates the speech parts when no other words are competing for the isolated phoneme, which indicates that the default integration system for auditory perceptual organization only applies to limited scenarios.

However, this project indeed has limitations. Due to the lack of evidence about the specific role of the competing word, the results cannot conclude whether the finding is generalizable to acoustically different word initial clusters. There are two major differences between Experiment 1 and Experiment 2: the presence of a competing word and the different phoneme combination conditions. I hypothesized that the presence of a competing word causes the drop in integration rate because, among the phoneme-combination conditions, two conditions (fricative+stop, fricative+nasal/liquid) had a high integration rate in Experiment 1. Therefore, for these two conditions, the only difference between Experiment 1 and Experiment 2 is the presence of a competing word. However, no available statistical evidence can support my hypothesis, so currently, I cannot preclude the possibility that the integration mechanism is not generalizable to acoustically different word initial clusters. Even though the results do not allow for generalization to more acoustically different clusters, given that those acoustically similar clusters all demonstrated a very high integration rate, those word pieces from Experiment 2 may still integrate when the competing word is not present. In addition, the lower integration rate for intact trials in Experiment 2 reflected a problem during stimuli creation. As described in the Results section, only stimulus pairs for split trials (a base word and a competing word) were recorded. Stimuli of intact trials were made by lateralizing the first phoneme of the competing words to the base words. Coarticulation made stimuli with stop, liquid, and nasal initials especially difficult for participants to hear. Even though in Experiment 2, participants still had a fair integration rate for intact trials, future experiments can avoid such a problem by recording all stimuli of intact and split trials.

For future research, the role of the competing word should be investigated explicitly by combining the stimuli of Experiment 1 and Experiment 2 so that researchers can determine whether the presence of a competing word is the cause of the low integration in Experiment 2. If a significant difference is found between the integration rate of these two sets of stimuli, they can conclude that the presence of a competing word has disrupted integration. Moreover, a different paradigm should be used to assess the generalizability of those acoustically different word initial clusters. The competing word is likely disrupting the integration because it is emanated at the same location as the isolated phoneme, and it is also lexically appropriate for the phoneme to attach to (they combine to create a real English word). The role of the competing word is probably related to the lexical effect, which is when the lexical knowledge affects listeners' perception of speech (Ganong, 1980). Therefore, when the competing word is no longer a full English word, the lexical effect may be eliminated. For example, instead of presenting a full word ("fact") as the competing word, future studies could only present a segment of the competing word /fa/, which reduces the lexical meaning of the competing word. With this new paradigm, I hypothesize that participants should have high integration rates across all acoustically different word initial clusters. Lastly, the lexical effect has probably played a role not only in competing for the isolated phoneme but also in successful integration processes since all integrated words are real English words. Future research could investigate the lexical effect on the default integration system (Freggens & Pitt, in preparation) by comparing the integration rate for nonwords and real words. If lexicality does affect integration, then it provides evidence that the default integration system operates on both bottom-up and top-down processing.

In summary, Freggens and Pitt (in preparation) proposed a default integration system that automatically integrates spatially separated speech parts, and the two experiments of this project assessed its generalizability. This project extends the literature of auditory perceptual organization, specifically how listeners organize spoken words in the environment, by providing further support and identifying limitations for this integration mechanism.

References

Barsz, K. (1991). Auditory pattern perception: The effect of tone location on the discrimination of tonal sequences. *Perception & Psychophysics*, *50*(3), 290–296. https://doi-org.proxy.lib.ohio-state.edu/10.3758/BF03206752

Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.41, retrieved 25 March 2021 from http://www.praat.org/

Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA, MIT Press.

Cutting, J. E. (1975). Aspects of phonological fusion. *Journal of Experimental Psychology. Human Perception & Performance*, *1*, 105–120. https://doi-org.proxy.lib.ohio-state.edu/10.1037/0096-1523.1.2.105

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, *83*(2), 114.

Cutting, J. E., & Day, R. S. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, *3*(2), 99–113. https://doi.org/10.1016/s0095-4470(19)31353-1

Denes, P. B. (1963). On the statistics of spoken English. *The Journal of the Acoustical Society of America*, *35*(6). https://doi.org/10.1524/stuf.1964.17.16.51

Freggens, M., & Pitt, M. A. (nd). Unpublished manuscript /"Manuscript in preparation"/

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic Properties of Stop Consonants. Journal of the Acoustical Society of America, 29, 107–116. https://doi-org.proxy.lib.ohio-state.edu/10.1121/1.1908634

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and Peripheral Channeling. *Music Perception*, *9*(2), 155–183. https://doi.org/10.2307/40285527

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252. https://doi.org/10.1121/1.1288413

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. *EURASIP Journal on Advances in Signal Processing*, *2009*(1). https://doi.org/10.1155/2009/298605

Köhler, W. (1947). Gestalt psychology: An introduction to new concepts in modern psychology. New York, Liveright Publishing Corporation.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, *80*, 479–483. https://doi-org.proxy.lib.ohio-state.edu/10.1121/1.394043

Noorden, van, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Technische Hogeschool Eindhoven. https://doi.org/10.6100/IR152538

Poltrock, S. E., & Hunt, E. (1977). Individual differences in phonological fusion and separation errors. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 62–74. https://doi.org/10.1037/0096-1523.3.1.62

Qualtrics. (2021). Qualtrics [Computer program]. Retrieved 25 March 2021 https://www.qualtrics.com/.

Sexton, M. A., & Geffen, G. (1981). Phonological fusion in dichotic monitoring. *Journal of experimental psychology. Human perception and performance*, *7*(2), 422–429. https://doi.org/10.1037//0096-1523.7.2.422

Appendix A

*Intact trials had silence instead of a first phoneme.

Phoneme-combination conditions	**Base**	**First phoneme**	**Integrated response**	**Segregated response**
FL	lag	/f/	flag	lag
FL	lash	/f/	flash	lash
FL	lease	/f/	fleece	*lease*
FL	lip	/f/	flip	lip
FL	loss	/f/	floss	loss
FR	rail	/f/	frail	rail
FR	rate	/f/	freight	*rate*
FR	rise	/f/	fries	*rise*
FR	risk	/f/	frisk	risk
FR	rose	/f/	froze	*rose*
SHR	rank	/sh/	shrank	rank
SHR	rue	/sh/	shrew	*rue*
SHR	rink	/sh/	shrink	rink
SHR	rug	/sh/	shrug	rug
SL	lash	/s/	slash	lash
SL	led	/s/	sled	led
SL	leap	/s/	sleep	*leap*
SM	myth	/s/	smith	*myth*
SM	mile	/s/	smile	mile
SN	nail	/s/	snail	nail
SN	nip	/s/	snip	nip
SN	news	/s/	snooze	*news*
SN	knows	/s/	snows	*knows*
SP	ban	/s/	span	*ban*
SP	beer	/s/	spear	*beer*
SP	bike	/s/	spike	*bike*
SP	bit	/s/	spit	*bit*
SP	bun	/s/	spun	*bun*
ST	damp	/s/	stamp	*damp*
ST	dart	/s/	start	*dart*
ST	dill	/s/	still	*dill*
ST	dove	/s/	stove	*dove*
ST	dump	/s/	stump	*dump*

SW	wag	/s/	swag	wag
SW	way	/s/	sway	way
SW	wheat	/s/	sweat	*wheat*
SW	wig	/s/	swig	wig

Appendix B

*Intact trials had silence instead of a competing word.

Phoneme-combination conditions	Base	Competing word	Integrated response	Segregated response
Fricative+Nasal/Liquid	rate	foil	freight	rate
FNL	rum	face	from	rum
FNL	wreak	shown	shriek	wreak
FNL	leak	sour	sleek	leak
FNL	leave	same	sleeve	leave
FNL	limb	sews	slim	limb
FNL	knees	sup	sneeze	knees
FNL	knob	sought	snob	knob
Fricative+Stop	ban	sold	span	ban
FS	bark	sat	spark	bark
FS	beer	sigh	spear	beer
FS	boil	seats	spoil	boil
FS	damp	seal	stamp	damp
FS	dart	sew	start	dart
FS	duck	sore	stuck	duck
FS	dump	sate	stump	dump
Fricative+Vowel	oak	feel	folk	oak
FV	owl	fame	foul	owl
FV	or	feet	four	or
FV	aim	seat	same	aim
FV	ease	sits	seize	ease
FV	ache	sheet	shake	ache
FV	aim	show	shame	aim
FV	eyes	sad	size	eyes
Nasal/Liquid+Vowel	ache	learn	lake	ache
NLV	eye	late	lie	eye
NLV	eat	mold	meet	eat
NLV	own	mark	moan	own
NLV	app	know	nap	app
NLV	odd	near	nod	odd
NLV	eyes	reach	rise	eyes
NLV	old	rat	rolled	old
Stop+Liquid	loan	bad	blown	loan
SL	lose	bake	blues	lose

SL	lamb	cold	clam	lamb
SL	lose	cage	clues	lose
SL	reef	gate	grief	reef
SL	rose	game	grows	rose
SL	rise	pose	prize	rise
SL	row	pin	pro	row
Stop+Vowel	add	beat	bad	add
SV	ear	buy	beer	ear
SV	air	dear	dare	air
SV	eye	date	die	eye
SV	add	pair	pad	add
SV	oak	peel	poke	oak
SV	ache	teach	take	ache
SV	aisle	tap	tile	aisle